Deep convolutional networks for automated detection of epileptogenic brain malformations

Gill RS [1*], Hong SJ [1], Fadaie F [1], Caldairou B [1], Bernhardt BC [1,2], Barba C [3], Brandt A [4], Coelho V [4], d'Incerti L [6], Lenge M [3], Semmelroch M [7], Bartolomei F [8], Cendes F [5], Deleo F [6], Guerrini R [3], Guye M [9], Jackson G [7], Schulze-Bonhage A [4], Mansi T [10], Bernasconi N [1], Bernasconi A [1]

[1] Neuroimaging of Epilepsy Laboratory, [2] Multimodal Imaging and Connectome Analysis Lab, Montreal Neurological Institute, McGill University, Montréal, QC, Canada, [3] Children's Hospital A. Meyer University of Florence, Florence, Italy, [4] Freiburg Epilepsy Center, University Medical Center, Freiburg, Germany, [5] University of Campinas, Campinas, Brazil, [6] Istituto Neurologico Carlo Besta, Milano, Italy, [7] The Florey Institute of Neuroscience and Mental Health and The University of Melbourne, Melbourne, VIC, Australia, [8] Aix Marseille Univ, INS, Marseille, France, [9] Aix Marseille Univ, CNRS, Marseille, France, [10] Siemens Medical Solutions, Medical Imaging Technologies, Princeton, NJ, USA

PURPOSE

Focal cortical dysplasia (FCD) is a surgically-amenable epileptogenic developmental malformation. On MRI, FCD typically presents with cortical thickening, hyperintensity, and blurring of the gray-white matter interface. These changes may be visible to the naked eye, or subtle and be easily overlooked [1]. Despite advances in MRI analytics, current surface-based algorithms [2-5] do not detect FCD in >50% of FCD lesions [6].

We propose a novel algorithm trained using a patch-based augmented dataset derived from patients with histologically-validated FCD, operating directly on MRI voxels, to distinguish the lesion from healthy tissue. Our method harnesses feature learning capability of convolutional neural networks (CNN). The algorithm was trained and cross-validated on multimodal MRI data from a single site (S1) and evaluated on independent data from S1 and six other sites worldwide (S2-S7) for a total of 107 subjects.

METHOD

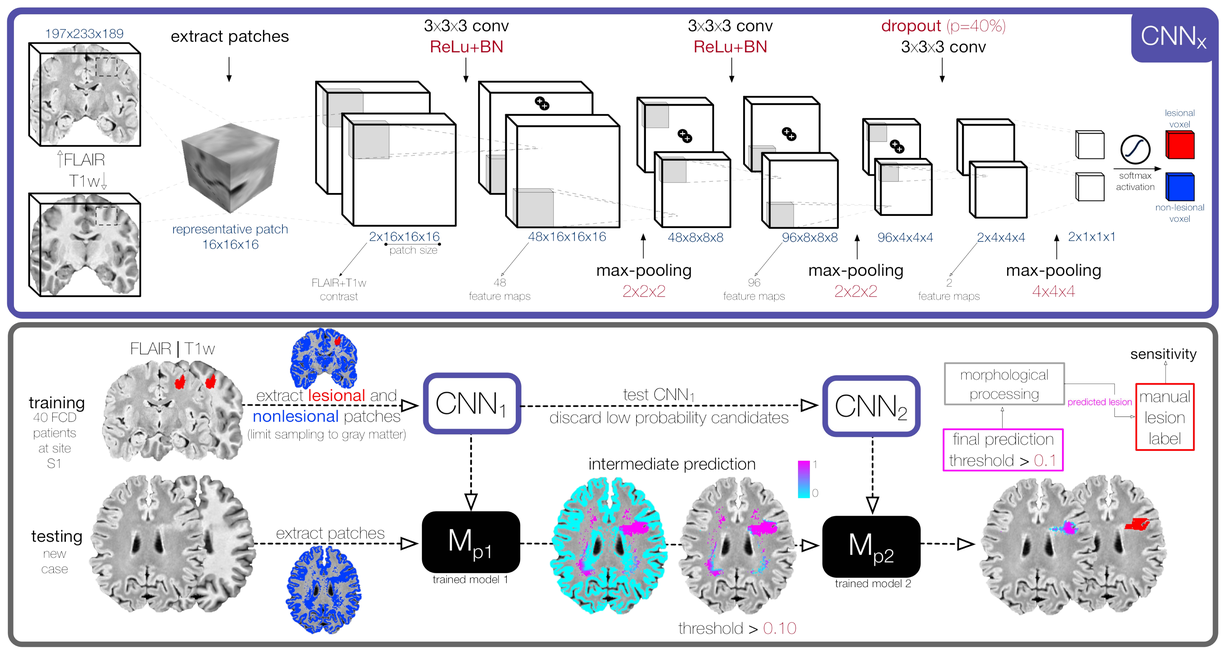

Figure 1. Upper panel: Convolutional network architecture (CNN_x) for two-label classification with three consecutive convolutions and max-pooling units, followed by a voxel-wise softmax classification using multimodal (FLAIR+T1w) patches. Each convolution is followed by rectified linear units (ReLu) to introduce non-linearity. Batch normalization (BN) and dropout serve as regularizers. Adadelta serves as the optimizer to minimize the binary crossentropy loss.
Lower panel: Training and testing schema using two-stage CNN_x cascade (CNN1/CNN2). To reduce training times, multimodal patches (2x16x16x16; centered around the voxel to classify) are sampled from a gray matter (GM) mask. This mask is generated using intra-subject z-score of FLAIR contrast, discarding hypointense voxels (z<0.1). Performance (sensitivity) is measured relative to the expert manual labels.

Classifier design. Volumetric datasets served as inputs to a two-stage cascaded CNN (**Figure 1**): the first CNN was designed to maximize sensitivity (*i.e.*, detecting a maximum number of lesional voxels), while the second optimized specificity (*i.e.*, reducing false positives).

Training. Volumetric 3T T1-weighted 3D-MPRAGE (T1w) and 3D-FLAIR MRI were collected in 40 patients (mean age: 28±9) with histologically-verified FCD. Routine MRI was initially reported as unremarkable in 80%. Images underwent intensity inhomogeneity correction and standardization. T1w images were registered linearly to the MNI152 template. FLAIR images were linearly mapped to T1w in MNI space.

Validation. At S1, a 5-fold cross validation repeated 20 times tested sensitivity (prediction co-localizing with manual FCD labels). Specificity was assessed by testing the model on 38 healthy controls (age: 30±7) and 63 temporal lobe epilepsy (TLE) patients as disease controls (age: 31±8). Sensitivity was tested in a separate cohort of 67 histologically-confirmed FCD (37 adults, age: 33±11; 30 children, age: 9±6) across sites with different scanners, field strengths, acquisition parameters and coils. We also compare the performance on the S1 dataset using a previously published method [4] based on an ensemble of RUSBoosted decision trees.

RESULTS

For S1, sensitivity was 87±4% (average of 35/40 FCD lesions detected with 2±1 extra-lesional clusters). Specificity was 95% in healthy controls (3±1 clusters in 2/38) and 90% in TLE (1±0 in 7/63).

For *cross-dataset classification* at 7 sites, overall sensitivity was 91% (61/67 lesions detected) with 3±2 extra-lesional clusters observed in 47/67 cases. Per-site sensitivity for S1-S7 was 100% (8/8 lesions detected, 2±2 extra-lesional clusters), 86% (17/19, 4±2), 89% (8/9, 2±1), 75% (6/8, 2±1), 100% (5/5, 5±2), 91% (10/11, 2±3), and 100% (7/7, 2±2), respectively. Stratifying patients based on age, sensitivity in children (2-18.5 years old) was 90% (27/30 FCD detected, 4±3 clusters) while in adults (>19 years old) it was 92% (34/37, 3±2). **Figure 2** shows test case examples.

Training and testing a surface-based classifier based on S1 dataset yielded a lower performance with a sensitivity of 83±2% (33/40 lesions detected), with 4±5 extra-lesional clusters. Specificity was 92% in healthy controls (1±0 cluster in 3/38).

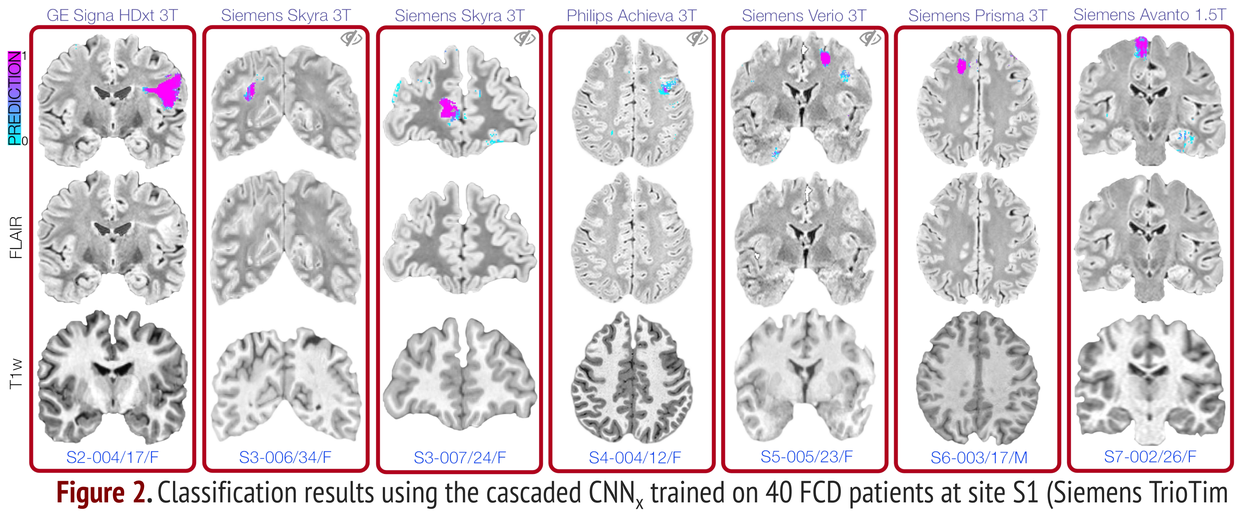

Figure 2. Classification results using the cascaded CNN_x trained on 40 FCD patients at site S1 (Siemens TrioTim 3T) to demonstrate generalizability for lesion detection along three axes of heterogeneity: scanner type, field strength (top labels), and age (bottom labels). The seven cases obtained using different scanners at six sites (excluding S1) are shown. The top row indicates the strength of prediction overlaid on the FLAIR, while the second/third rows show the corresponding FLAIR and T1w, respectively. The bottom labels are read as site-patient-ID/age/gender. MRI-negative cases are identified with .

CONCLUSION

We present the first deep learning method to segment FCD, with multicentric validation. Operating on routine multi-contrast MRI in voxel-space, our algorithm provides the highest performance to date. Furthermore, we demonstrated generalizability of a model trained on a single-site dataset by showing robust performance across independent cohorts from various centers worldwide with different age, scanner hardware and sequence parameters. Notably, ~50% of FCD lesions were missed by conventional MRI visual inspection. Easy implementation, minimal pre-processing, performance gains, and inference time of <6 minutes/case make this classifier the ideal platform for large-scale clinical use, particularly in "MRI-negative" FCD.

REFERENCES

[**1**] Bernasconi A, et al. Advances in MRI for 'cryptogenic' epilepsies. Nat Rev Neurol 2011, [**2**] Adler S., et al. Novel surface features for automated detection of focal cortical dysplasias in paediatric epilepsy. NeuroImage Clin 2017, [**3**] Hong, S.-J. et al. Automated detection of cortical dysplasia type II in MRI-negative epilepsy. Neurology 83, 48–55 (2014), [**4**] Gill RS, et al. Automated detection of epileptogenic cortical malformations using multimodal MRI. DLMIA/ML-CDS, MICCAI Proceedings (eds. Cardoso MJ, et al.) 349–356 (2017), [**5**] Tan YL, et al. Quantitative surface analysis of combined MRI and PET enhances detection of focal cortical dysplasias. Neuroimage 166, 10–18 (2017), [**6**] Kini, LG, et al. Computational analysis in epilepsy neuroimaging: A survey of features and methods. Neuroimage: Clin 11, 515–529 (2016)

neuroimaging of epilepsy laboratory noel

*ravnoor.gill@mail.mcgill.ca
www.bic.mni.mcgill.ca/~noel